THE EXACT HORIZON-TWO QUERIED-GRADIENT WORST CASE OF NESTEROV'S FAST GRADIENT METHOD IS ONE-DIMENSIONAL

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. Let W_N be the worst-case value of $\min_{0\le k\le N}\|\nabla f(x_k)\|^2$ over L-smooth convex functions with $\|x_0 - x_\star\| \le R$, where $x_0, \dots, x_N$ are the points at which Nesterov's fast gradient method queries the gradient. Du (arXiv:2608.26719) proves $W_N = L^2R^2/S_N$ for every $N \ge 7$ in every dimension $d \ge N-4$, with $S_N = \sum_{k\le N} t_k^2$, and $W_1 = L^2R^2/4$, and for $2 \le N \le 6$ prints floating-point values while explicitly claiming no closed form, no uniqueness and no rank pattern. We settle the first of these horizons. Writing a for the coefficient of $\nabla f(x_1)$ in $x_0 - x_2$ (for Nesterov's method $a = 1 + (t_1-1)/t_2$), we show that for every real $a \in [5/4, 13/10]$ the horizon-two worst case of the two-step scheme with coefficient a equals $L^2R^2/(2+a)^2$ exactly, is attained by a one-dimensional (rank-one) instance, and is certified by a dual multiplier vector that is a rational function of a. In particular $W_2 = L^2R^2/\big(3 + \frac{t_1-1}{t_2}\big)^2 = 0.0928513193\ldots \cdot L^2R^2$, strictly below the Kim–Fessler bound L^2R^2/S_2. At horizon one we also give the exact post-gradient value $1/9$, a number Taylor, Hendrickx and Glineur printed as $LR/3.00$ from a semidefinite solve and for which no certificate had been stated. Every theorem is stated over the finite interpolation data and is machine-checked in Lean 4 with Mathlib; no step rests on floating-point arithmetic or on an exhaustive search. The transfer to the function class, which rests on the cited interpolation theorem or on the source's explicit projection-envelope interpolant, is recorded separately and is not part of the formal development.

1. Introduction

The objects. Fix $L > 0$, $R \ge 0$ and $d \ge 1$, and let $\mathcal{F}_{0,L}(\mathbb{R}^d)$ be the class of convex differentiable functions $f\colon \mathbb{R}^d \to \mathbb{R}$ whose gradient is L-Lipschitz. Nesterov's fast gradient method (FGM) [6], in the normalisation of [1, equations (2)–(5)], is

$$
\begin{aligned}
y_0 &= x_0, \qquad t_0 = 1, \quad y_{k+1} = x_k - \tfrac{1}{L}\nabla f(x_k),\\
t_{k+1} &= \tfrac{1+\sqrt{1+4t_k^2}}{2}, \qquad x_{k+1} = y_{k+1} + \tfrac{t_k-1}{t_{k+1}}(y_{k+1} - y_k).
\end{aligned}
\tag{1}
$$

The *horizon* N counts update steps: the method queries gradients at the $N+1$ points $x_0, \dots, x_N$ and, if one more gradient step is taken, produces the post-gradient points $y_0, \dots, y_{N+1}$. With $S_N := \sum_{k=0}^N t_k^2$, the two performance measures of [1] are

$$W_N(L,R,d) := \sup_{f,\,x_0,\,x_\star}\ \min_{0\le k\le N}\|\nabla f(x_k)\|^2, \tag{2}$$

$$Y_N(L,R,d) := \sup_{f,\,x_0,\,x_\star}\ \min_{0\le k\le N+1}\|\nabla f(y_k)\|^2, \tag{3}$$

the suprema being over $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$, $x_\star \in \arg\min f$ and x_0 with $\|x_0 - x_\star\| \le R$. We call W_N the *queried* and Y_N the *post-gradient* measure. Both scale as L^2R^2 (Section 2), so nothing is lost by taking $L = R = 1$.

What the source proves, and where it stops. The dimension-free upper bound $W_N \le L^2R^2/S_N$ is due to Kim and Fessler [2, Theorem 6, inequality (5.7)], whose Table 3 also reports tight numerical values of $LR/\min_{i\le N}\|\nabla f(x_i)\|$ for FGM at selected horizons (2.0, 3.3, 5.9, 13.8, ... at $N = 1, 2, 4, 10, \dots$, to one decimal place). Du [1] proves that the bound is exact for every $N \ge 7$ and every dimension $d \ge N-4$ ([1, Theorem 7.2], by an explicit $(N-4)$-dimensional adversary) and settles $N = 1$ by hand ([1, Proposition 8.1]: $W_1 = L^2R^2/4$). For the five horizons $2 \le N \le 6$ the source prints, in its Table 1, the floating-point outputs of an exact-interpolation semidefinite program ($\widehat{W}_2 = 0.0928513212$ with primal–dual gap 3.4×10^{-9}) and is careful about their status:

> "The horizons $2 \le N \le 6$ do not form a proved continuation of either Proposition 8.1 or the lifted $N \ge 7$ branch. [...] We claim no closed forms, optimizer uniqueness, or rank pattern for those five numerical optima." [1, Section 8]

and, of the table itself, "Neither the numerical objective nor the reported gap is a rigorous enclosure or certificate, and signs at or below this numerical scale have no mathematical significance." For the post-gradient measure the source prints floats at *every* horizon (its Table 2, $\widehat{Y}_1 = 0.1111111112$) and states that "no analytical formula is claimed; neither the numerical dual objectives nor the reported gaps are rigorous enclosures."

What this paper settles. We give the exact value at the first exceptional horizon, and we give it for a whole interval of extrapolation coefficients. At horizon two the momentum sequence enters the iteration only through the single number

$$a_{\mathrm{FGM}} := 1 + \frac{t_1 - 1}{t_2} = 1 + \frac{\sqrt{5}-1}{1+\sqrt{7+2\sqrt{5}}} = 1.28175352\ldots, \tag{4}$$

the coefficient of $\nabla f(x_1)$ in $x_0 - x_2$ (Section 2). Consider the two-step scheme $x_1 = x_0 - \nabla f(x_0)$, $x_2 = x_1 - a\nabla f(x_1)$ with an arbitrary real coefficient a. Our main results, stated over the finite interpolation data in Section 4, are:

- for every $a \in [5/4, 13/10]$ the horizon-two worst case of $\min_{k\le 2}\|\nabla f(x_k)\|^2$ is exactly $1/(2+a)^2$, certified by a dual multiplier vector whose entries are rational functions of a and attained by a *rank-one* Gram matrix, i.e. by a one-dimensional instance in which all three queried gradients equal $(x_0 - x_\star)/(2+a)$ (Theorem 4.1);
- consequently, for Nesterov's own coefficient,

$$W_2 = \frac{L^2R^2}{\left(3 + \frac{t_1-1}{t_2}\right)^2}, \tag{5}$$

  where $1/(2+a_{\mathrm{FGM}})^2 = 0.0928513193\ldots$ lies in $(0.09285131, 0.09285133)$, strictly below the Kim–Fessler value $1/S_2 = 0.11862966\ldots$ (Theorem 4.3, Corollary 4.4, Proposition 4.5);
- at horizon one the exact post-gradient value is $Y_1 = L^2R^2/9$, with an explicit rational certificate (Proposition 3.2). The *number* is not new: Taylor, Hendrickx and Glineur [3, Table 4] print $LR/3.00$ for exactly this quantity from a semidefinite solve, with no exactness claim. The exact statement and its certificate are.

The source's own horizon-one result $W_1 = L^2R^2/4$ is reproduced with its certificate as a positive control (Proposition 3.1), and a set of negative controls (Propositions 3.3 and 4.6) records that the hypotheses are load-bearing and the values are sharp on both sides.

The rank-one finding is the part the source explicitly declines to claim. It is not an instance of a general pattern: for the closely related smallest-gradient criterion of plain gradient descent on smooth weakly convex functions, Rotaru, Glineur and Patrinos [4] exhibit worst-case instances that are one-, two- or three-dimensional depending on the constant step size (Section 6), so the dimension of a worst case for a smallest-gradient measure is a fact about the method, its parameters and the horizon, not about the measure. At $N = 5, 6$ the numerical optimum for FGM is not rank one (Section 5, unverified numerics), and the source's own extremal instances for $N \ge 7$ are $(N-4)$-dimensional.

Method. Everything is proved as a statement about the finite interpolation data of the points $x_0, \ldots, x_N, x_\star$: a Gram matrix, function values, the initial-distance constraint and the interpolation inequalities of [3]. Upper bounds are exact dual certificates — nonnegative multipliers, a convex combination of the performance constraints, and an explicit $LDL^\top$ factorisation of the dual slack — found by solving a semidefinite program numerically, reading off the active set, and then solving the Karush–Kuhn–Tucker system exactly, over $\mathbb{Q}$ at horizon one and over the rational function field $\mathbb{Q}(a)$

at horizon two. Lower bounds are explicit rank-one data, realised by the one-dimensional projection-envelope (Huber) function that already appears in [1, Proposition 8.1] and in classical constructions going back to Drori and Teboulle [5]. Every theorem below, with the sole exception of the transfer to the function class (Corollary 4.7), is machine-checked in Lean 4 with Mathlib [7] (Section 7); the formal development contains no floating-point computation and no exhaustive search, and its only numerical content is a handful of rational brackets on $\sqrt{5}$ and $\sqrt{7+2\sqrt{5}}$.

2. Preliminaries

2.1. **Scaling.** If $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $\|x_0 - x_\star\| \le R > 0$, then $\tilde f(x) := f(Rx)/(LR^2)$ lies in $\mathcal{F}_{0,1}(\mathbb{R}^d)$, the FGM iterates of $\tilde f$ from x_0/R are x_k/R, and the squared gradient norms of $\tilde f$ at x_k/R are those of f at x_k divided by L^2R^2. Hence $W_N(L,R,d) = L^2R^2\, W_N(1,1,d)$ and likewise for Y_N; the case $R = 0$ is trivial. From now on $L = R = 1$, and we write $g_k := \nabla f(x_k)$.

2.2. **Interpolation data.** For a finite index set I, a family of triples $(x_i, g_i, f_i)_{i\in I}$ in $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$ is $\mathcal{F}_{0,1}$-*interpolable* if some $f \in \mathcal{F}_{0,1}(\mathbb{R}^d)$ has $f(x_i) = f_i$ and $\nabla f(x_i) = g_i$ for all i. The interpolation theorem of Taylor, Hendrickx and Glineur [3, Theorem 4], specialised to $\mu = 0$ and $L = 1$, says that this holds if and only if

$$A_{ij} := f_i - f_j - \langle g_j, x_i - x_j\rangle - \tfrac{1}{2}\|g_i - g_j\|^2 \;\ge\; 0 \qquad \text{for all } i \ne j \text{ in } I. \tag{6}$$

Only the easy direction is used for the upper bounds in this paper: every $f \in \mathcal{F}_{0,1}(\mathbb{R}^d)$ satisfies (6) at any finite set of points. The hard direction (sufficiency) is what identifies the values of the finite problems below with the suprema (2)–(3); it is cited, never reproved, and in every case treated here the extremal data are in addition realised by an explicit function, so the transfer does not depend on it.

For a first-order method the points are affine combinations of $x_0 - x_\star$ and the gradients, and (6) becomes affine in the function values and in the entries of the Gram matrix of the vectors $(x_0 - x_\star, g_0, g_1, \dots)$. We always take $x_\star$ as the origin of the Gram matrix, with $g_\star = 0$ and $f_\star = 0$. A Gram matrix is nothing but a positive semidefinite matrix, and we state positive semidefiniteness of a matrix $P = (p_{ij})$ as nonnegativity of its quadratic form, $c^\top Pc \ge 0$ for all real c. Since every positive semidefinite $n \times n$ matrix is the Gram matrix of n vectors in $\mathbb{R}^n$, the problems below are the *dimension-unrestricted* performance estimation problems, exactly the objects the tables of [1] and [2] approximate.

2.3. **The three finite problems.**

Definition 2.1 (horizon one, queried)**.** Let $I = \{0, 1, \star\}$ with $x_1 = x_0 - g_0$, and let $P = (p_{ij})_{0\le i,j\le 2}$ be the Gram matrix of $(v_0, v_1, v_2) = (x_0 - x_\star, g_0, g_1)$. The data (P, f_0, f_1) are *feasible* if $P \succeq 0$, $p_{00} \le 1$, and the six inequalities (6) hold. Let $\mathcal{W}_1 \subset \mathbb{R}$ be the set of values $\min(p_{11}, p_{22})$ over feasible data.

Definition 2.2 (horizon one, post-gradient)**.** Let the points be x_0, $x_1 = y_1 = x_0 - g_0$, $y_2 = x_1 - g_1$ and $x_\star$, with gradients $g_0, g_1, h_2 := \nabla f(y_2)$, and let $P = (p_{ij})_{0\le i,j\le 3}$ be the Gram matrix of $(x_0 - x_\star, g_0, g_1, h_2)$. (Since $t_0 = 1$, $y_1 = x_1$; so Y_1 measures the three points $y_0 = x_0$, $y_1 = x_1$ and y_2.) Feasibility means $P \succeq 0$, $p_{00} \le 1$ and the twelve inequalities (6); $\mathcal{Y}_1 \subset \mathbb{R}$ is the set of values $\min(p_{11}, p_{22}, p_{33})$ over feasible data.

Definition 2.3 (horizon two with coefficient a)**.** Fix $a \in \mathbb{R}$. Let the points be x_0, $x_1 = x_0 - g_0$, $x_2 = x_0 - g_0 - a\, g_1$ and $x_\star$, and let $P = (p_{ij})_{0\le i,j\le 3}$ be the Gram matrix of $(x_0 - x_\star, g_0, g_1, g_2)$. Feasibility means $P \succeq 0$, $p_{00} \le 1$ and the twelve inequalities (6) (which now involve a); $\mathcal{W}_2(a) \subset \mathbb{R}$ is the set of values $\min(p_{11}, p_{22}, p_{33})$ over feasible data.

For Nesterov's method (1) at $L = 1$ one has $x_1 = y_1 = x_0 - g_0$, $y_2 = x_1 - g_1$ and $x_2 = y_2 + \frac{t_1 - 1}{t_2}(y_2 - y_1) = x_0 - g_0 - \big(1 + \frac{t_1-1}{t_2}\big) g_1$, so the horizon-two problem of FGM is Definition 2.3 with $a = a_{\mathrm{FGM}}$ as in (4). Note that $t_1 = (1+\sqrt{5})/2$ and, since $1 + 4t_1^2 = 7 + 2\sqrt{5}$, $t_2 = (1 + \sqrt{7 + 2\sqrt{5}})/2$.

By the easy direction of (6), for every d,

$$W_1(1,1,d) \le \sup \mathcal{W}_1, \qquad Y_1(1,1,d) \le \sup \mathcal{Y}_1, \qquad W_2(1,1,d) \le \sup \mathcal{W}_2(a_{\mathrm{FGM}});$$

by the interpolation theorem the three inequalities are equalities once d is at least the number of Gram vectors. The theorems below concern the sets $\mathcal{W}_1$, $\mathcal{Y}_1$ and $\mathcal{W}_2(a)$; the transfer to (2)–(3) is recorded separately (Corollary 4.7).

2.4. **Certificates.** All upper bounds have the same shape. For a feasible datum with metric value $m = \min_k p_{kk}$ (over the measured indices k), an identity of the form

$$\nu - m = \nu\,(1 - p_{00}) + \sum_{i\neq j} \lambda_{ij} A_{ij} + \sum_k \mu_k\,(p_{kk} - m) + \sum_r d_r\,(w_r^\top P w_r), \tag{7}$$

valid for *all* real (P, f, m) as a polynomial identity, with $\lambda_{ij} \ge 0$, $\mu_k \ge 0$, $\sum_k \mu_k = 1$ and $d_r \ge 0$, proves $m \le \nu$ on feasible data, because every term on the right is then nonnegative. (The coefficient of m forces $\sum_k \mu_k = 1$, and the function values must cancel, which is the usual balance condition on λ.) The quadratic part $\sum_r d_r w_r w_r^\top$ is the $LDL^\top$ factorisation of the dual slack matrix.

3. Horizon one

The two horizon-one results are not new mathematics and are included as calibration: the first is [1, Proposition 8.1], certificate included; the second is a number published as data in [3] and as a float in [1].

Proposition 3.1 ($W_1 = 1/4$; [1, Proposition 8.1])**.** *Every feasible datum of Definition 2.1 satisfies* $\min(p_{11}, p_{22}) \le 1/4$*, and* $1/4$ *is the greatest element of* $\mathcal{W}_1$*. The maximum is attained by the rank-one data* $x_0 - x_\star = e$*,* $g_0 = g_1 = e/2$ *for a unit vector* e*, i.e.* $p_{00} = 1$*,* $p_{01} = p_{02} = 1/2$*,* $p_{11} = p_{12} = p_{22} = 1/4$*, with* $(f_0, f_1) = (3/8, 1/8)$*.*

Proof. The identity (7) holds with $\nu = 1/4$, $\mu = (0,1)$, $\lambda = 1$ on the four pairs $(0,1), (1,0), (1,\star), (\star,1)$ and 0 elsewhere, and the single square $\frac14\|v_0 - 2v_2\|^2$, $v_0 = x_0 - x_\star$, $v_2 = g_1$:

$$\tfrac14 - m = \tfrac14(1 - p_{00}) + (A_{01} + A_{10} + A_{1\star} + A_{\star 1}) + (p_{22} - m) + \tfrac14\,(p_{00} - 4p_{02} + 4p_{22}).$$

This is literally the argument of [1, Proposition 8.1]: $A_{01} + A_{10}$ and $A_{1\star} + A_{\star 1}$ are the two cocoercivity inequalities, and the square is Cauchy–Schwarz. The data listed are feasible (the Gram form is $(c_0 + \frac12 c_1 + \frac12 c_2)^2 \ge 0$, and every A_{ij} is a nonnegative rational) and have value 1/4. □

Proposition 3.2 ($Y_1 = 1/9$)**.** *Every feasible datum of Definition 2.2 satisfies* $\min(p_{11}, p_{22}, p_{33}) \le 1/9$*, and* $1/9$ *is the greatest element of* $\mathcal{Y}_1$*. The maximum is attained by the rank-one data* $x_0 - x_\star = e$*,* $g_0 = g_1 = h_2 = e/3$*, i.e.* $p_{00} = 1$*,* $p_{0k} = 1/3$ *and* $p_{kl} = 1/9$ *for* $k, l \ge 1$*, with* $(f_0, f_1, f_2) = (5/18, 1/6, 1/18)$*.*

Proof. Index the points $0, 1, 2, \star$ for $x_0, x_1, y_2, x_\star$ and the Gram vectors $v_0, \dots, v_3$ for $x_0 - x_\star, g_0, g_1, h_2$. The identity (7) holds with $\nu = 1/9$, $\mu = (0, \frac{1}{12}, \frac{11}{12})$ on (p_{11}, p_{22}, p_{33}),

$$\lambda_{01} = \tfrac13, \quad \lambda_{02} = \tfrac{1}{12}, \quad \lambda_{10} = \tfrac16, \quad \lambda_{12} = 1, \quad \lambda_{20} = \tfrac{1}{12}, \quad \lambda_{21} = \tfrac{7}{12},$$
$$\lambda_{2\star} = \tfrac23, \quad \lambda_{\star 0} = \tfrac16, \quad \lambda_{\star 1} = \tfrac14, \quad \lambda_{\star 2} = \tfrac14,$$

$\lambda_{0\star} = \lambda_{1\star} = 0$, and the three squares

$$\tfrac{1}{576}\|8v_0 - 6v_1 - 9v_2 - 9v_3\|^2 + \tfrac{1}{960}\|10v_1 - 9v_2 - v_3\|^2 + \tfrac{11}{40}\|v_2 - v_3\|^2.$$

The identity was checked in exact rational arithmetic; the formal proof instantiates the Gram form at the three vectors and closes the inequality by linear arithmetic. The listed data are feasible — the Gram form is $(c_0 + \frac13(c_1 + c_2 + c_3))^2$, and the twelve A_{ij} are nonnegative rationals, all zero except $A_{0\star} = 2/9$ and $A_{1\star} = 1/9$ — and have value 1/9. □

The extremal instance is the one-dimensional Huber function $\phi(x) = \max\{gx - g^2/2 : 0 \le g \le 1/3\}$, whose gradient clips x to $[0, 1/3]$. From $x_0 = 1$ the three points 1, 2/3 and 1/3 all lie in the saturated region, the last one exactly on the kink. The same template with $K = [0, 1/2]$ gives Proposition 3.1; both are instances of the projection-envelope function ϕ_K of [1, Lemma 3.1].

Proposition 3.3 (controls at horizon one)**.** (i) $1/5$ *is not an upper bound for* $\mathcal{W}_1$*, and* $1/10$ *is not an upper bound for* $\mathcal{Y}_1$*.*

(ii) *The initial-distance hypothesis is load-bearing: the data of Proposition 3.1 scaled by two* ($p_{00} = 4$*,* $p_{01} = p_{02} = 2$*,* $p_{11} = p_{12} = p_{22} = 1$*,* $f = (3/2, 1/2)$) *satisfy* $P \succeq 0$ *and all six inequalities* (6) *and have value* $1 > 1/4$*.*

(iii) *The measure is load-bearing: every element of* $\mathcal{Y}_1$ *is strictly less than* $1/4$*.*

Proof. (i) and (iii) are immediate from the two propositions; (ii) is a direct check of the scaled data. □

4. Horizon two

4.1. **The one-parameter family.** Throughout, $s := 1/(2 + a)$.

Theorem 4.1 (exact value for every coefficient in $[5/4, 13/10]$)**.** *Let* $a \in [5/4, 13/10]$ *be real. Every feasible datum of Definition 2.3 satisfies*

$$\min(p_{11}, p_{22}, p_{33}) \le \frac{1}{(2 + a)^2},$$

and $1/(2 + a)^2$ *is the greatest element of* $\mathcal{W}_2(a)$*. The maximum is attained by the rank-one data* $x_0 - x_\star = e$*,* $g_0 = g_1 = g_2 = s\,e$ *for a unit vector* e *— that is,* $p_{00} = 1$*,* $p_{0k} = s$*,* $p_{kl} = s^2$ *for* $k, l \ge 1$ *— with function values* $(f_0, f_1, f_2) = \big(s - \frac{s^2}{2},\ s - \frac{3s^2}{2},\ \frac{s^2}{2}\big)$*.*

The upper bound is an identity (7) whose coefficients are rational functions of a; we state it in full, since the certificate is the theorem.

Proof of the upper bound. Index the points $0, 1, 2, \star$ and write v_0, v_1, v_2, v_3 for the Gram vectors $x_0-x_\star$, g_0, g_1, g_2. Take $\nu = 1/(2 + a)^2$, take $\mu = (0, \frac{1}{12}, \frac{11}{12})$ on (p_{11}, p_{22}, p_{33}), take the multipliers

$$\lambda_{01} = \frac{8a^2 - 6a + 4}{12(2 + a)}, \qquad \lambda_{02} = \frac{20 - 8a - 3a^2}{12(2 + a)}, \qquad \lambda_{10} = \frac{5a^2 - 9a + 10}{12(2 + a)}, \qquad \lambda_{12} = \frac{5 + 3a}{12},$$
$$\lambda_{20} = \frac{1}{12}, \qquad \lambda_{21} = \frac{5}{12}, \qquad \lambda_{2\star} = \frac{2}{2 + a}, \qquad \lambda_{\star 0} = \frac{2 - a}{2(2 + a)},$$
$$\lambda_{\star 1} = \lambda_{\star 2} = \frac{1}{4}, \qquad \lambda_{0\star} = \lambda_{1\star} = 0,$$

and the dual slack $S(a) = \sum_{r=1}^{3} d_r w_r w_r^\top$ with

$$d_1 = \frac{1}{(2 + a)^2}, \qquad w_1 = \Big(1,\ \tfrac{a^2}{4} - 1,\ -\tfrac{(2+a)^2}{8},\ -\tfrac{(2+a)^2}{8}\Big),$$
$$d_2 = \frac{(2 - a)(4 + a^2)}{16(2 + a)}, \qquad w_2 = \Big(0,\ 1,\ \tfrac{3a^3-18a^2+28a-40}{6(2-a)(4+a^2)},\ \tfrac{(2+a)(3a^2-4)}{6(2-a)(4+a^2)}\Big),$$
$$d_3 = \frac{q(a)}{72\,(16 - a^4)}, \qquad w_3 = (0,\ 0,\ 1,\ -1),$$

where $q(a) := 33a^4 - 132a^3 + 280a^2 - 656a + 592$. With these values, (7) is an identity in $\mathbb{Q}(a)[p_{ij}, f_i, m]$. Multiplying through by $D(a) := 576\,(2+a)(16-a^4) = 576\,(2+a)^2(2-a)(4+a^2)$ turns every coefficient into a polynomial in a with integer coefficients, of degree at most 8; this cleared identity is what the formal proof checks, as a single polynomial identity in the thirteen data variables and a.

It remains to see that every multiplier is nonnegative on $[5/4, 13/10]$. The factors $2 + a$, $2 - a$, $4 + a^2$, $16 - a^4 = (2 - a)(2 + a)(4 + a^2)$ and $5 + 3a$ are positive there; $8a^2 - 6a + 4$ and $5a^2 - 9a + 10$ are positive everywhere (discriminants -92 and -119); $20 - 8a - 3a^2$ is decreasing for $a > 0$ and equals 4.53 at $a = 13/10$, so it is positive on the interval (the formal proof uses $(a - \frac{5}{4})(\frac{13}{10} - a) \ge 0$ instead); and the quartic q is positive on the interval, its Bernstein coefficients on $[5/4, 13/10]$ being 32.25, 28.29, 24.37, 20.49, 16.65 (to two decimals), all positive. Hence $d_1, d_2, d_3 \ge 0$ and $\lambda_{ij} \ge 0$, and (7) gives $m \le \nu$. □

Proof of attainment. Let $s = 1/(2+a)$, so $0 < s < 1/2$ for $a > 0$. The Gram form of the listed data is $(c_0 + s(c_1 + c_2 + c_3))^2 \geq 0$, $p_{00} = 1$, and a direct computation using $sa = 1 - 2s$ shows that every A_{ij} vanishes except $A_{0\star} = s(1-s)$ and $A_{1\star} = s(1-2s)$, both positive. The value is $\min(s^2, s^2, s^2) = 1/(2+a)^2$. (Attainment therefore holds for every $a > 0$; the interval enters only through the upper bound.) □

The extremal data are the constant-gradient trajectory of the one-dimensional Huber function $\phi_s(x) = \max_{g\in[0,s]}\{gx - g^2/2\}$, $\nabla\phi_s(x) = \min(\max(x, 0), s)$, started at $x_0 = 1$ with $x_\star = 0$: the queried points are 1, $1-s$ and $1-(1+a)s = s$, the last exactly on the kink, so all three gradients equal s. Equivalently s is the largest constant gradient for which the three queried iterates of the two-step scheme stay in the saturated region. Complementary slackness is visible in the certificate: the two slack inequalities $A_{0\star}, A_{1\star}$ are exactly the two whose multipliers vanish.

Remark 4.2. Nothing is claimed for a outside $[5/4, 13/10]$. The interval is a rational bracket of a_{FGM} chosen so that the positivity conditions above are elementary; the certificate itself is a rational function of a and the identity holds for all a, so the same argument extends to any a at which all multipliers are nonnegative, but that extension has not been carried out.

4.2. **Nesterov's coefficient.**

Theorem 4.3 (the exact horizon-two value of FGM)**.** *Let* $t_1 = (1+\sqrt{5})/2$, $t_2 = (1+\sqrt{7+2\sqrt{5}})/2$ *and* $a_{\mathrm{FGM}} = 1 + (t_1 - 1)/t_2$. *Then* $a_{\mathrm{FGM}} \in [5/4, 13/10]$, *and the greatest element of* $\mathcal{W}_2(a_{\mathrm{FGM}})$ *is*

$$\frac{1}{(2+a_{\mathrm{FGM}})^2} = \frac{1}{\left(3 + \frac{t_1-1}{t_2}\right)^2} = \frac{1}{\left(3 + \frac{\sqrt{5}-1}{1+\sqrt{7+2\sqrt{5}}}\right)^2}.$$

Proof. From $2.2360679 < \sqrt{5} < 2.2360680$ one gets $1.61803395 < t_1 < 1.6180340$ and $3.3870540 < \sqrt{7+2\sqrt{5}} < 3.3870543$, hence $2.1935270 < t_2 < 2.19352715$ and $1.28175349 \leq a_{\mathrm{FGM}} \leq 1.28175355$, so $a_{\mathrm{FGM}} \in [5/4, 13/10]$. Apply Theorem 4.1. (Since $t_1^2 = t_1 + 1$, one also has $(t_1 - 1)/t_2 = 1/(t_1 t_2)$, an elementary rewriting not used further.) □

Corollary 4.4 (a rigorous enclosure)**.** $0.09285131 < \dfrac{1}{(2+a_{\mathrm{FGM}})^2} < 0.09285133.$

Proof. From the brackets on a_{FGM} in the previous proof, $10.7699059 \leq (2+a_{\mathrm{FGM}})^2 \leq 10.7699064$, and the reciprocals of these bounds lie in the stated interval. □

The source's printed $\widehat{W}_2 = 0.0928513212$ lies inside this enclosure; the exact value $0.0928513193\ldots$ differs from it by 1.8×10^{-9}, within the source's own reported primal–dual gap of 3.4×10^{-9}. The reciprocal square root of the exact value is $2 + a_{\mathrm{FGM}} = 3.28175\ldots$, which is the "3.3" of [2, Table 3] at $N = 2$.

Proposition 4.5 (strictly below the Kim–Fessler bound)**.** *Every element* v *of* $\mathcal{W}_2(a_{\mathrm{FGM}})$ *satisfies* $v < 1/S_2$, *where* $S_2 = 1 + t_1^2 + t_2^2$.

Proof. By Theorem 4.3 and Corollary 4.4, $v < 0.09285133$; from the brackets on t_1, t_2, $S_2 < 8.43$, so $1/S_2 > 0.1186$. □

This is the sign of the source's numerical $\Delta_2 = \widehat{W}_2 - 1/S_2 = -2.58 \times 10^{-2}$ [1, Table 1] as a theorem: the dimension-free relaxation of [2] is not exact at horizon two. Numerically $1/S_2 = 0.11862966\ldots$

Proposition 4.6 (controls at horizon two)**.** (i) $9/100$ *is not an upper bound for* $\mathcal{W}_2(a_{\mathrm{FGM}})$; $\mathcal{W}_2(a_{\mathrm{FGM}})$ *is nonempty.*

(ii) *The extrapolation coefficient is load-bearing:* $1/(2+a_{\mathrm{FGM}})^2 \neq 1/(2+\frac{13}{10})^2$, *so by Theorem 4.1 the two-step scheme with coefficient* $13/10$ *has a different exact horizon-two value from Nesterov's method.*

Proof. (i) is Corollary 4.4 together with attainment. For (ii), $1/(2+13/10)^2 = 100/1089$, which is $0.09182\ldots < 0.09285131$. □

4.3. **Transfer to the function class.**

Corollary 4.7 (not part of the formal development). *For every $d \geq 1$, $W_1(1,1,d) = 1/4$, $Y_1(1,1,d) = 1/9$ and $W_2(1,1,d) = 1/(2+a_{\mathrm{FGM}})^2$; more generally the horizon-two worst case of the two-step scheme with coefficient $a \in [5/4, 13/10]$ equals $1/(2+a)^2$ in every dimension. With the scaling of Section 2 this is* (5).

Proof. Upper bounds: the data of any $f \in \mathcal{F}_{0,1}(\mathbb{R}^d)$ at the relevant points satisfy (6), hence are feasible, so Propositions 3.1, 3.2 and Theorem 4.1 apply. Lower bounds: the extremal data are realised in $\mathbb{R}^1$ (and hence, along any line, in $\mathbb{R}^d$) by the projection-envelope functions ϕ_K with $K = [0, 1/2]$, $[0, 1/3]$ and $[0, 1/(2+a)]$ respectively, which belong to $\mathcal{F}_{0,1}$ with $\nabla\phi_K = \mathrm{Proj}_K$ by [1, Lemma 3.1], and whose iterates were computed above. Alternatively the sufficiency half of [3, Theorem 4] transfers any feasible datum. Neither the membership $\phi_K \in \mathcal{F}_{0,1}$ nor the interpolation theorem is formalised; this corollary is the only statement in the paper that rests on a cited result. □

5. Beyond horizon two: what is not settled

The following observations come from an interior-point solver written for this project and from the closed forms it suggested; none of them is part of the formal development, and none is claimed as a theorem.

Remark 5.1 (numerical observations, unverified). Rebuilding the dimension-unrestricted queried-gradient program (the analogue of Definition 2.3 at horizon N) for $N \leq 7$ reproduces the source's Table 1 to 7–9 significant digits. The optimal Gram matrix returned by the solver has rank one at $N = 1, 2, 3, 4$ (rank-one residual 5.4×10^{-12} at $N = 4$) and does *not* at $N = 5, 6, 7$ (residuals of order 10^{-2}). Reading the rank-one optimum as a one-dimensional instance gives, at $N = 4$, the Huber function with saturation level $c = 0.2066931631$: the queried points saturate until $x_3 = c$ lands on the kink, and x_4 overshoots into the linear region where $g_4 = x_4 < 0$. Running FGM on a constant gradient s from $x_0 = 1$ gives $y_k = 1 - \eta_k s$ and $x_k = 1 - \beta_k s$ with

$$\eta_0 = \beta_0 = 0, \qquad \eta_{k+1} = \beta_k + 1, \qquad \beta_{k+1} = \eta_{k+1} + \tfrac{t_k - 1}{t_{k+1}}(\eta_{k+1} - \eta_k),$$

so $\eta_3 = 2 + a_{\mathrm{FGM}}$ recovers Theorem 4.3. The same ansatz predicts $1/\eta_4^2 = 0.042722063689\ldots$ at $N = 3$ (against the source's 0.0427220648, and against the solver's 0.0427220636883) and, at $N = 4$, where the saturated ansatz is no longer optimal, $(1 - \beta_4/\eta_4)^2 = 0.029184501651\ldots$ (source: 0.0291845018). For the post-gradient measure the source's $\widehat{Y}_2 = 0.0545452170$ is not $3/55 = 0.05454545\ldots$, so the clean value at $N = 1$ does not continue in the obvious way.

Why the formal development stops at horizon two is a matter of cost, not of mathematics. At horizon two the momentum sequence enters through one coefficient, the certificate lives in $\mathbb{Q}(a)$, and the cleared identity has thirteen data variables and degree 8 in a. At horizon three there are three independent coefficients (those of g_1 in $x_0 - x_2$ and in $x_0 - x_3$, and of g_2 in $x_0 - x_3$ — algebraically dependent as numbers, but independent as certificate parameters), the dual would have to be solved over a multivariate rational function field, and the identity grows to nineteen data variables and degree about 12. At horizons five and six the numerical worst case is not one-dimensional, so a different ansatz is needed before any certificate can be sought. The rank-one lower bound, by contrast, is cheap at every horizon; it is tight at $N = 2$ (Theorem 4.3), numerically tight at $N = 3$, and not tight at $N = 4, 5, 6$ (Remark 5.1).

6. Sources and prior numerics

The source. All quotations from [1] are from the version-1 e-print (the only version on arXiv on 2026-09-03; primary category math.OC, no journal reference), compiled here with `tectonic`; equation, theorem and table numbers are read from the compiled auxiliary file: FGM is (2)–(5), W_N is (7), the Kim–Fessler bound is Theorem 2.1, the projection-envelope lemma is Lemma 3.1, the main theorem is Theorem 7.2, the first horizon is Proposition 8.1, and the two numerical tables are Tables 1 and 2 of Section 8. The source's ancillary Python file `anc/fgm_best_queried_gradient_pep.py` fixes the normalisation and the two measures used here.

Kim and Fessler. The bibliography entry `KimFessler2018` of [1] names [2], that is arXiv:1607.06764 (SIAM J. Optim. 28(2), 2018), *not* the same authors' JOTA paper arXiv:1803.06600 on optimising the efficiency of first-order methods for decreasing the gradient, which has no Table 3 and no smallest-gradient-over-iterates criterion. In the arXiv version 4 of [2], read here as the PDF, the chain $\min_{i\le N+1}\|\nabla f(y_i)\| \le \min_{i\le N}\|\nabla f(x_i)\| \le LR/\sqrt{S_N}$ (with $y_{N+1} = x_N - \frac{1}{L}\nabla f(x_N)$) is Theorem 6, inequality (5.7), in Section 5.2 ("A worst-case bound for the gradient norm of FGM"), and Table 3, in Section 6.2, is captioned "Tight worst-case bounds on $\frac{LR}{\min_{i\in\{0,\ldots,N\}}\|\nabla f(x_i)\|}$, the reciprocal of the gradient norm, of GM, FGM, OGM, […]"; its FGM column reads 2.0, 3.3, 5.9, 13.8, 32.8, 56.4, 83.6, 104.7, 114.2 at $N = 1, 2, 4, 10, 20, 30, 40, 47, 50$, so $N = 3, 5, 6$ are not computed there and no closed form is stated. The source [1, Theorem 2.1] cites the same chain as "Theorem 5.2, equation (5.7), in Section 5.2 of the published version" of [2]; the published text is behind the publisher's paywall and was not consulted, so every Kim–Fessler number in this paper is the arXiv version-4 number.

Taylor, Hendrickx and Glineur. In the version-6 e-print of [3], compiled here, the interpolation theorem is Theorem 4 (Section 2) and the table of gradient-norm numerics for FGM is Table 4 in Section 4.3, captioned "FGM and MFGM: comparison between theoretical bounds and numerical results for the criteria $\|\nabla f(y_N)\|_2$(last) and $\min_i\|\nabla f(y_i)\|_2$ (best). Results obtained with [Mosek]" (their typographical emphasis on y is not reproduced). Its "FGM, best" column reads $LR/3.00$ at their $N = 2$ and $LR/5.84$ at their $N = 4$. Their $N = 2$ covers y_0, y_1, y_2, which is the Y_1 of (3), and $(LR/3.00)^2 = L^2R^2/9$; the alignment is confirmed by their $N = 4$, where $1/5.84^2 = 0.02932\ldots$ against the source's $\widehat{Y}_3 = 0.0293398935$. Thus 1/9 was published as a three-significant-figure solver output in the version-6 e-print of [3] (the copy compiled here is byte-identical to the live arXiv e-print; the journal version, Math. Program. 161 (2017) 307–345, is confirmed by Crossref), with no exactness claim, certificate or extremal instance, which is why Proposition 3.2 is presented as a verification of published data rather than as a new value.

Rotaru, Glineur and Patrinos. The e-print of [4] (version 2, 22 January 2026, read from the live e-print; its sources do not compile under `tectonic`, so no statement numbers are cited) was read for one fact only: its constant-stepsize theorem for smooth weakly convex functions bounds $\min_{0\le i\le N}\|\nabla f(x_i)\|^2$ for gradient descent, and its summary table of tightness proofs lists, for that class, one-dimensional worst-case functions on two step-size ranges, two-dimensional worst-case functions or interpolation data on two others, and three-dimensional interpolation data on the remaining range. (Version 1 of the e-print stated the three-dimensional case as a conjecture; version 2 states it as a proposition with a proof. Their normalisation is by $f(x_0) - f_\star$ rather than by $\|x_0 - x_\star\|$.)

Searches. The decimal strings 0.0928513 and 0427220648 occur in a local text corpus of arXiv sources only in [1] itself; the value $0.0928513\ldots$ matches no entry of the OEIS; the source has no citing works recorded in OpenAlex as of 2026-09-03. Each negative is read as "not found", not as "new".

7. Verification

Every theorem, proposition and corollary of Sections 3 and 4, with the sole exception of Corollary 4.7, has been formally verified in Lean 4 with Mathlib, in the directory `FgmQueriedSmall` of the `LeanProblemSpec` development, in a repository that is private at the time of writing (repository reference to be added at submission). The development is five modules (`Pep`, `Cert1`, `Cert2`, `Cert3`, `Horizon2`) and sixteen headline declarations, whose exact statements are reproduced in Appendix A. It contains no `sorry` and no declared axiom, and every one of the sixteen declarations depends on exactly the three axioms `propext`, `Classical.choice` and `Quot.sound`, as reported by `#print axioms`; `native_decide` is not used anywhere, so no statement rests on compiled evaluation, and there is no enumeration or search. Positive semidefiniteness is stated as nonnegativity of the quadratic form for all real coefficient vectors, which is what a Gram matrix satisfies and is exactly what the certificates instantiate. The upper bounds are the cleared identities of Section 2.4, checked by the `ring` normaliser; the positivity conditions are discharged by `nlinarith` and `positivity`; the decimal literals in Corollary 4.4 are exact rationals; the square roots enter only through the brackets stated in the proof of Theorem 4.3, each proved from $\sqrt{x}^2 = x$. The one measurement worth recording is the cost of the horizon-two identity: asked to verify it in one piece, `ring` ran for more than forty minutes of wall time at 5.4 GB and was stopped; pre-expanding each of the sixteen terms of the certificate (three squares, ten interpolation multipliers, the radius term and two performance terms) into a flat linear form in the data variables (so that the normaliser never multiplies two large polynomials), giving the thirteen squares and interpolation terms a lemma each in the modules `Cert1`–`Cert3`, and assembling all sixteen in `Horizon2`, brings the whole horizon-two bill to about seven processor-minutes (34, 16, 10 and 351 seconds for `Cert1`, `Cert2`, `Cert3` and `Horizon2`; the horizon-one module `Pep` costs a further 9.7 seconds), the split being a pure cost split with no mathematics in the boundary. All of it builds under Lean `v4.32.0` (toolchain `leanprover/lean4`) with Mathlib pinned at its `v4.32.0` tag.

References

[1] Y. Du, *When a Relaxed PEP Is Exact: The Sharp Queried-Gradient Rate of Nesterov's Fast Gradient Method*, arXiv:2608.26719 (27 August 2026), primary category `math.OC`. Version 1 is the only version and carries no journal reference on its arXiv abstract page as of 2026-09-03. All numbering here is from the e-print of version 1 compiled with `tectonic`.

[2] D. Kim and J. A. Fessler, *Generalizing the optimized gradient method for smooth convex minimization*, SIAM J. Optim. 28 (2018), no. 2, 1920–1950, `doi:10.1137/17M112124X`; arXiv:1607.06764. Journal data confirmed by Crossref and the arXiv abstract page; result numbers from the arXiv version-4 PDF (the published numbering was not consulted).

[3] A. B. Taylor, J. M. Hendrickx and F. Glineur, *Smooth strongly convex interpolation and exact worst-case performance of first-order methods*, Math. Program. 161 (2017), no. 1–2, 307–345, `doi:10.1007/s10107-016-1009-3` (Crossref; published online 17 May 2016); arXiv:1502.05666. Result numbers from the arXiv version-6 e-print compiled with `tectonic`.

[4] T. Rotaru, F. Glineur and P. Patrinos, *Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions*, arXiv:2406.17506 (version 2, 22 January 2026); Math. Program., published online 26 May 2026, `doi:10.1007/s10107-025-02313-1` (Crossref; no volume or page numbers assigned, and no journal reference on the arXiv abstract page, as of 2026-09-03).

[5] Y. Drori and M. Teboulle, *Performance of first-order methods for smooth convex minimization: a novel approach*, Math. Program. 145 (2014), no. 1–2, 451–482, `doi:10.1007/s10107-013-0653-0` (Crossref); arXiv:1206.3209.

[6] Yu. E. Nesterov, *A method of solving a convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR 269 (1983), 543–547; English translation: Soviet Math. Dokl. 27 (1983), 372–376 (Zbl 0535.90071).

[7] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381, `doi:10.1145/3372885.3373824`.

Appendix A. The formal statements

The sixteen declarations below are reproduced from the Lean sources exactly, except that long lines are broken at whitespace to fit the page (Lean is whitespace-insensitive inside a statement) and proofs are omitted. The Gram entries are `p`ij, the function values `f`i, and the quantifier over `c0 c1 …` is the positive-semidefiniteness hypothesis. Module `Pep`:

```
def FeasW1 (p00 p01 p02 p11 p12 p22 f0 f1 : ℝ) : Prop :=
  (∀ c0 c1 c2 : ℝ, 0 ≤ c0*c0*p00 + 2*c0*c1*p01 + 2*c0*c2*p02 + c1*c1*p11 +
      2*c1*c2*p12 + c2*c2*p22) ∧
  p00 ≤ 1 ∧
  0 ≤ - (1/2)*p11 - (1/2)*p22 + f0 - f1 ∧
  0 ≤ - (1/2)*p11 + f0 ∧
  0 ≤ (1/2)*p11 + p12 - (1/2)*p22 - f0 + f1 ∧
  0 ≤ - (1/2)*p22 + f1 ∧
  0 ≤ p01 - (1/2)*p11 - f0 ∧
  0 ≤ p02 - p12 - (1/2)*p22 - f1

def PepW1 : Set ℝ := { v | ∃ p00 p01 p02 p11 p12 p22 f0 f1 : ℝ,
    FeasW1 p00 p01 p02 p11 p12 p22 f0 f1 ∧ v = min p11 p22 }

theorem pepW1_le (p00 p01 p02 p11 p12 p22 f0 f1 : ℝ)
    (h : FeasW1 p00 p01 p02 p11 p12 p22 f0 f1) :
    min p11 p22 ≤ (1/4 : ℝ)

theorem pepW1_isGreatest : IsGreatest PepW1 (1/4 : ℝ)
```

```
def FeasY1 (p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 : ℝ) : Prop :=
  (∀ c0 c1 c2 c3 : ℝ, 0 ≤ c0*c0*p00 + 2*c0*c1*p01 + 2*c0*c2*p02 + 2*c0*c3*p03 +
      c1*c1*p11 + 2*c1*c2*p12 + 2*c1*c3*p13 + c2*c2*p22 + 2*c2*c3*p23 + c3*c3*p33) ∧
  p00 ≤ 1 ∧
  0 ≤ - (1/2)*p11 - (1/2)*p22 + f0 - f1 ∧
  0 ≤ - (1/2)*p11 - p23 - (1/2)*p33 + f0 - f2 ∧
  0 ≤ - (1/2)*p11 + f0 ∧
  0 ≤ (1/2)*p11 + p12 - (1/2)*p22 - f0 + f1 ∧
  0 ≤ - (1/2)*p22 - (1/2)*p33 + f1 - f2 ∧
  0 ≤ - (1/2)*p22 + f1 ∧
  0 ≤ (1/2)*p11 + p12 + p13 - (1/2)*p33 - f0 + f2 ∧
  0 ≤ (1/2)*p22 + p23 - (1/2)*p33 - f1 + f2 ∧
  0 ≤ - (1/2)*p33 + f2 ∧
  0 ≤ p01 - (1/2)*p11 - f0 ∧
  0 ≤ p02 - p12 - (1/2)*p22 - f1 ∧
  0 ≤ p03 - p13 - p23 - (1/2)*p33 - f2

def PepY1 : Set ℝ := { v | ∃ p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 : ℝ,
    FeasY1 p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 ∧
    v = min p11 (min p22 p33) }

theorem pepY1_le (p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 : ℝ)
    (h : FeasY1 p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2) :
    min p11 (min p22 p33) ≤ (1/9 : ℝ)

theorem pepY1_isGreatest : IsGreatest PepY1 (1/9 : ℝ)

theorem pepW1_not_le_one_fifth : ¬ (∀ v ∈ PepW1, v ≤ (1:ℝ)/5)

theorem pepY1_not_le_one_tenth : ¬ (∀ v ∈ PepY1, v ≤ (1:ℝ)/10)

theorem pepW1_radius_needed :
    ¬ (∀ p00 p01 p02 p11 p12 p22 f0 f1 : ℝ,
        (∀ c0 c1 c2 : ℝ, 0 ≤ c0*c0*p00 + 2*c0*c1*p01 + 2*c0*c2*p02 + c1*c1*p11 +
            2*c1*c2*p12 + c2*c2*p22) →
        0 ≤ - (1/2)*p11 - (1/2)*p22 + f0 - f1 →
        0 ≤ - (1/2)*p11 + f0 →
        0 ≤ (1/2)*p11 + p12 - (1/2)*p22 - f0 + f1 →
        0 ≤ - (1/2)*p22 + f1 →
        0 ≤ p01 - (1/2)*p11 - f0 →
        0 ≤ p02 - p12 - (1/2)*p22 - f1 →
        min p11 p22 ≤ (1:ℝ)/4)

theorem pepY1_lt_pepW1_value : ∀ v ∈ PepY1, v < (1:ℝ)/4
```

Module `Cert1` (the predicate; the thirteen certificate-term lemmas of `Cert1`–`Cert3` are internal and not listed):

```
def FeasW2 (a p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 : ℝ) : Prop :=
  (∀ c0 c1 c2 c3 : ℝ, 0 ≤ c0*c0*p00 + 2*c0*c1*p01 + 2*c0*c2*p02 + 2*c0*c3*p03 +
      c1*c1*p11 + 2*c1*c2*p12 + 2*c1*c3*p13 + c2*c2*p22 + 2*c2*c3*p23 + c3*c3*p33) ∧
  p00 ≤ 1 ∧
  0 ≤ - (1/2)*p11 - (1/2)*p22 + f0 - f1 ∧
  0 ≤ - (1/2)*p11 + ((-1)*a)*p23 - (1/2)*p33 + f0 - f2 ∧
  0 ≤ - (1/2)*p11 + f0 ∧
  0 ≤ (1/2)*p11 + p12 - (1/2)*p22 - f0 + f1 ∧
  0 ≤ - (1/2)*p22 + (1 + (-1)*a)*p23 - (1/2)*p33 + f1 - f2 ∧
  0 ≤ - (1/2)*p22 + f1 ∧
  0 ≤ (1/2)*p11 + (1*a)*p12 + p13 - (1/2)*p33 - f0 + f2 ∧
  0 ≤ ((-1/2) + 1*a)*p22 + p23 - (1/2)*p33 - f1 + f2 ∧
  0 ≤ - (1/2)*p33 + f2 ∧
  0 ≤ p01 - (1/2)*p11 - f0 ∧
  0 ≤ p02 - p12 - (1/2)*p22 - f1 ∧
  0 ≤ p03 - p13 + ((-1)*a)*p23 - (1/2)*p33 - f2

def PepW2 (a : ℝ) : Set ℝ :=
  { v | ∃ p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 : ℝ,
    FeasW2 a p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 ∧
    v = min p11 (min p22 p33) }
```

Module `Horizon2`:

```
theorem pepW2_le (a p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2 : ℝ)
    (ha1 : (5:ℝ)/4 ≤ a) (ha2 : a ≤ 13/10)
    (h : FeasW2 a p00 p01 p02 p03 p11 p12 p13 p22 p23 p33 f0 f1 f2) :
    min p11 (min p22 p33) ≤ 1 / (2 + a)^2

theorem pepW2_isGreatest (a : ℝ) (ha1 : (5:ℝ)/4 ≤ a) (ha2 : a ≤ 13/10) :
```

```
    IsGreatest (PepW2 a) (1 / (2 + a)^2)

noncomputable def tSeq : ℕ → ℝ
  | 0 => 1
  | (k + 1) => (1 + Real.sqrt (1 + 4 * (tSeq k) ^ 2)) / 2

noncomputable def fgmCoeff2 : ℝ := 1 + (tSeq 1 - 1) / tSeq 2

theorem fgm_pepW2_isGreatest :
    IsGreatest (PepW2 fgmCoeff2) (1 / (2 + fgmCoeff2) ^ 2)

theorem fgm_W2_enclosure :
    (0.09285131 : ℝ) < 1 / (2 + fgmCoeff2) ^ 2 ∧ 1 / (2 + fgmCoeff2) ^ 2 < 0.09285133

theorem pepW2_not_le_nine_hundredths : ¬ (∀ v ∈ PepW2 fgmCoeff2, v ≤ (9:ℝ)/100)

theorem pepW2_lt_kimFessler :
    ∀ v ∈ PepW2 fgmCoeff2, v < 1 / (1 + tSeq 1 ^ 2 + tSeq 2 ^ 2)

theorem pepW2_stepsize_matters :
    1 / (2 + fgmCoeff2) ^ 2 ≠ 1 / (2 + (13:ℝ)/10) ^ 2

theorem pepW2_nonempty : (PepW2 fgmCoeff2).Nonempty
```

Correspondence with the text:

- Proposition 3.1: `pepW1_le`, `pepW1_isGreatest`; Proposition 3.2: `pepY1_le`, `pepY1_isGreatest`;
- Proposition 3.3: the four controls of module `Pep`;
- Theorem 4.1: `pepW2_le`, `pepW2_isGreatest`; Theorem 4.3: `fgm_pepW2_isGreatest`; Corollary 4.4: `fgm_W2_enclosure`; Proposition 4.5: `pepW2_lt_kimFessler`;
- Proposition 4.6: `pepW2_not_le_nine_hundredths`, `pepW2_stepsize_matters`, `pepW2_nonempty`.

The witness data are the lemmas `feasW1_witness`, `feasY1_witness` and `feasW2_witness`; the last is stated for every $a > 0$ and s with $s(2 + a) = 1$.